Bias–Variance and Cross-Validation

Nipun Batra and teaching staff

IIT Gandhinagar

August 14, 2025

Table of Contents

Roadmap

What is the Bias–Variance Tradeoff?

Important: The Core Dilemma

Simpler model $\Rightarrow$ misses structure (High Bias)
Complex model $\Rightarrow$ fits noise (High Variance)

Key Points

Today: Intuition $\rightarrow$ Math $\rightarrow$ Practice

A Real-World Analogy: Weather Prediction

Example: Simple Model: “Tomorrow = Today”

High Bias, Low Variance

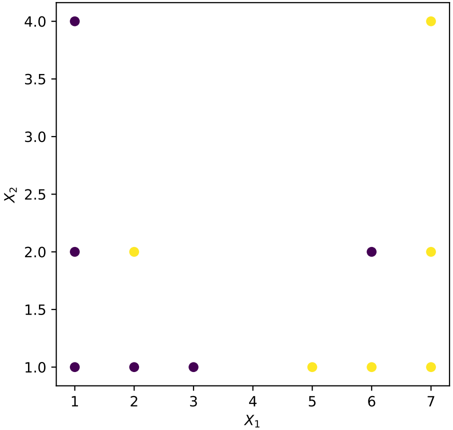
Example: Huge Model: 1000+ features

Low Bias, High Variance

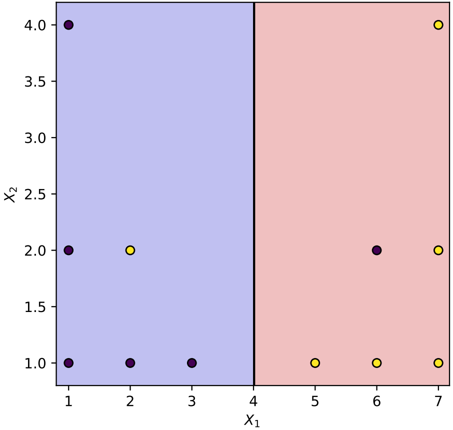
Aim: Find the Goldilocks zone (just right).

A Question!

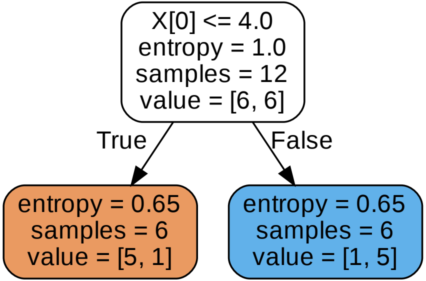
What would be the decision boundary of a decision tree classifier?



Decision Boundary vs Tree (depth = 1)

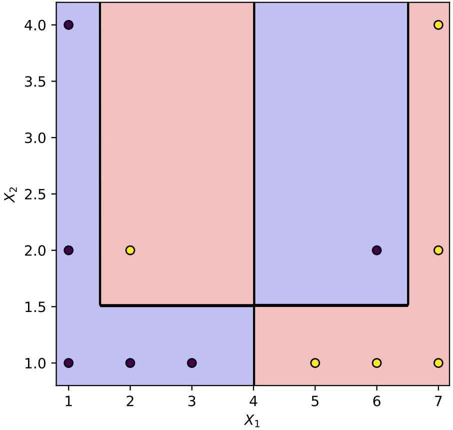


Decision Boundary

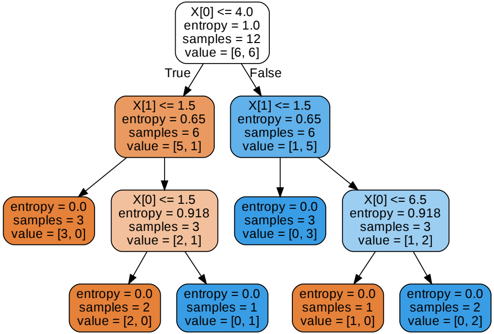


Decision Tree

Decision Boundary vs Tree (no depth limit)



Decision Boundary



Decision Tree

Are deeper trees always better?

Deeper trees learn more complex boundaries.

Are deeper trees always better?

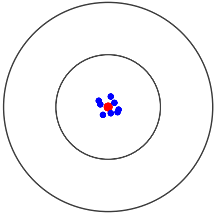
But they can generalize poorly (overfit).

Three concepts from what we saw

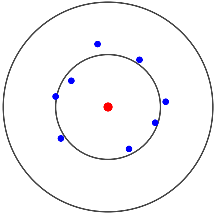
Key Points

1. **Bias**: Error from restrictive assumptions
2. **Variance**: Sensitivity to data fluctuations
3. **Irreducible Error**: Data noise

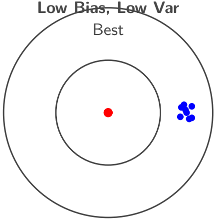
Dartboard Analogy: Four Scenarios



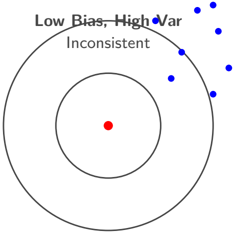
Low Bias, Low Var
Best



Low Bias, High Var
Inconsistent



High Bias, Low Var
Consistent & wrong



High Bias, High Var
Worst

Bias–Variance Decomposition

Definition: Fundamental Equation

$$\mathbb{E}\left[(Y - \hat{f}(X))^2\right] = \underbrace{(\mathbb{E}[\hat{f}(X)] - f(X))^2}_{\text{Bias}^2} + \underbrace{\mathbb{V}[\hat{f}(X)]}_{\text{Variance}} + \underbrace{\mathbb{V}[\varepsilon]}_{\text{Irreducible noise}}$$

Intuitions

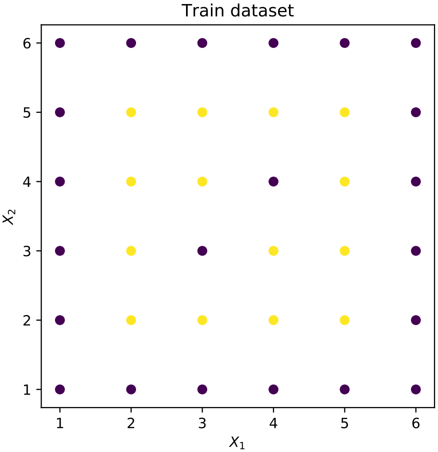
Example: Bias

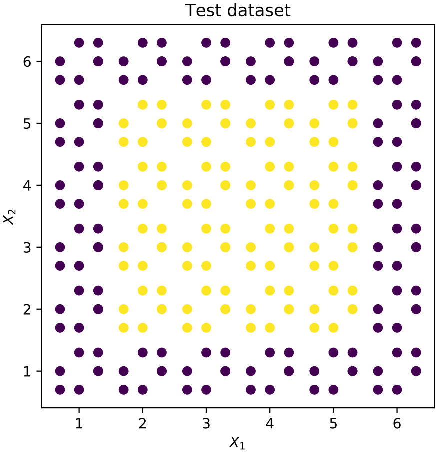
Average model $\neq$ truth (e.g., linear on curved data).

Example: Variance

Model changes a lot across different train sets.

An example: Train vs Test



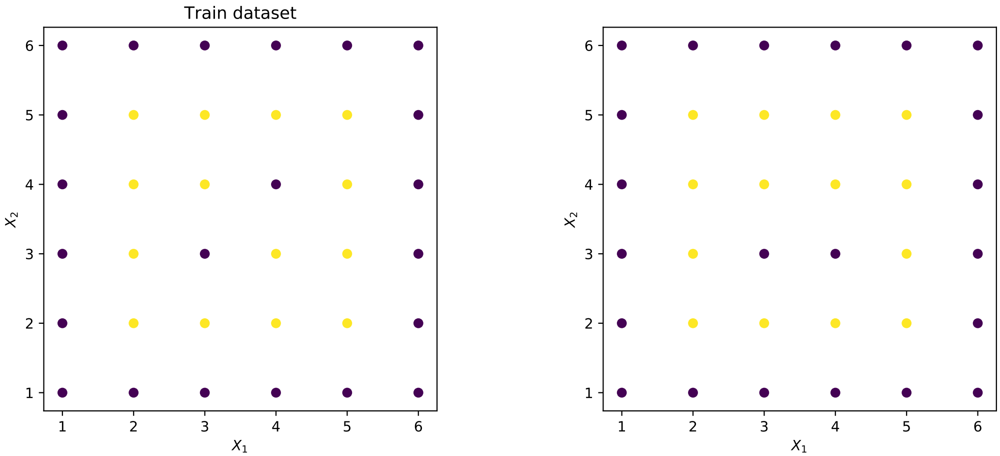


Train

Test

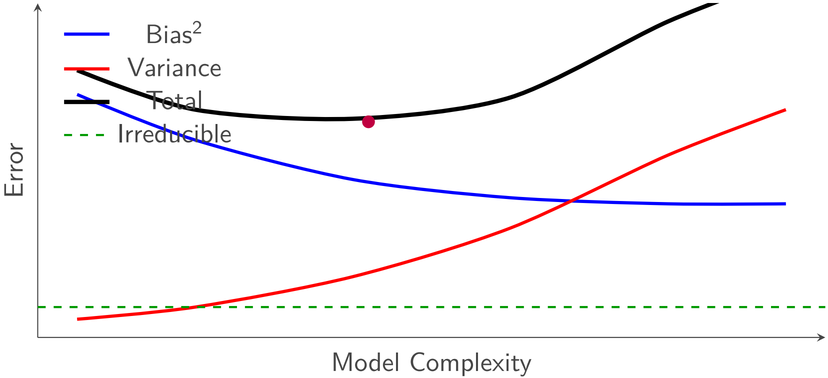
Intuition for Variance

Small data changes $\Rightarrow$ very different models.

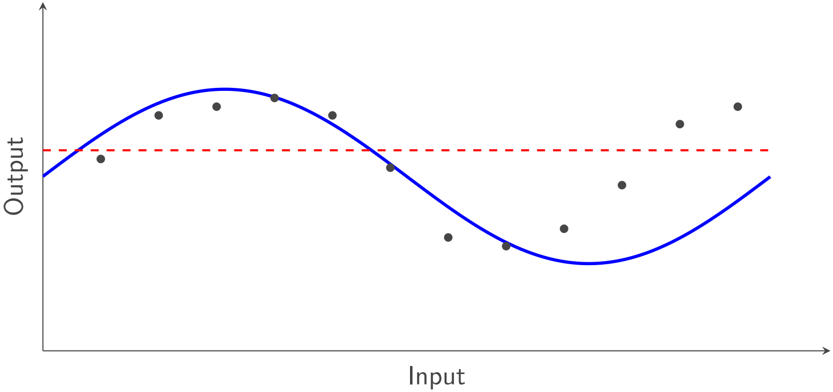


Dataset 1 Dataset 2

Bias–Variance vs Complexity (schematic)



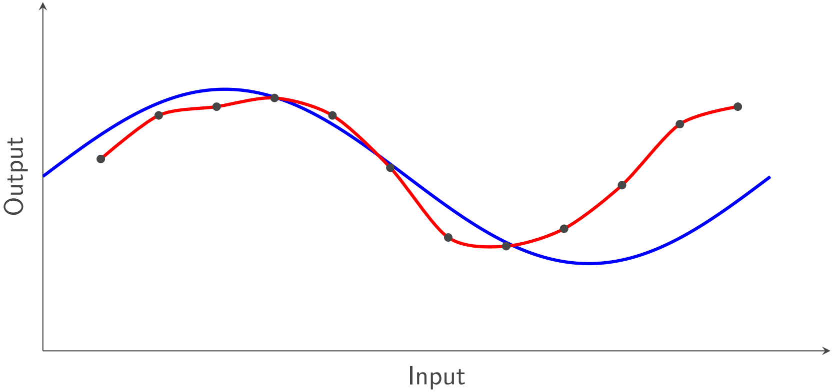
Underfitting (too simple)



Important: High Bias

Systematic error; both train/test errors high.

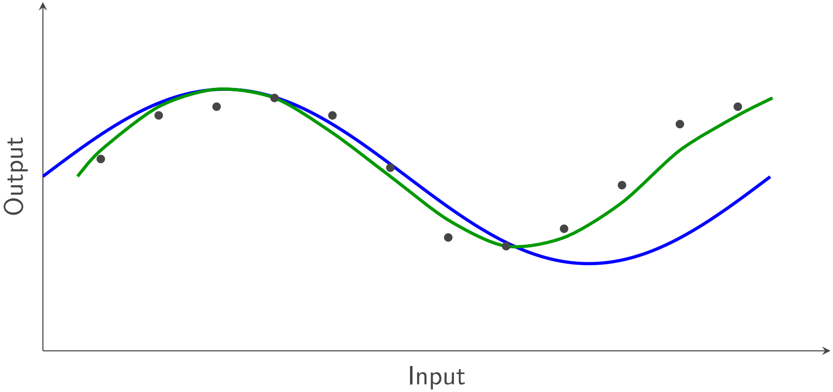
Overfitting (too complex)



Important: High Variance

Memorizes noise; train error low, test error high.

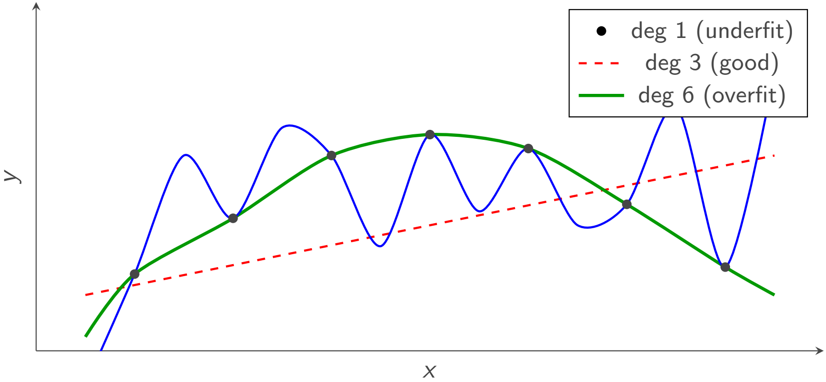
Good Fit (sweet spot)



Example: Goldilocks

Balanced bias and variance ⇒ best generalization.

Interactive: Polynomial Degrees



Roadmap

Why training error fails for model selection

Example: Optimistic bias

Training error $\downarrow$ as capacity $\uparrow$, even if test error $\uparrow$.

Key Points

We need an unbiased estimate of generalization.

Cross-Validation: Core Idea

Definition: Philosophy

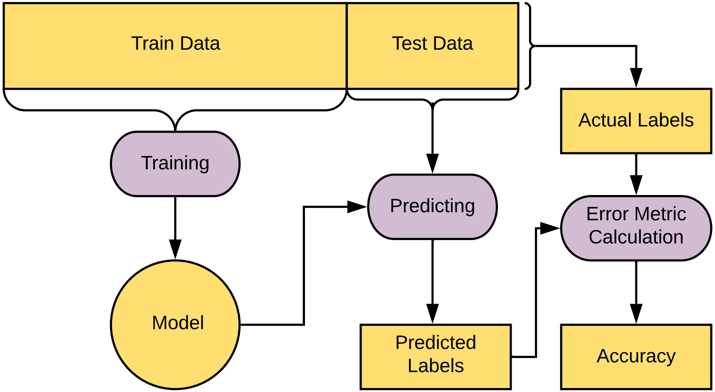
Split into k folds; train on $k-1$, validate on 1; rotate; average.

Benefits of Cross-Validation

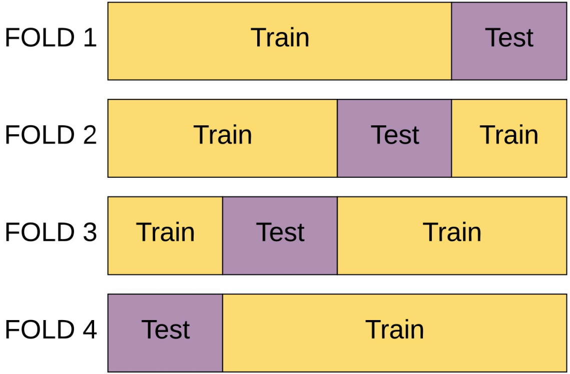
Key Points

- Uses all data for both train and validation
- Less sensitive to a single split
- Honest model comparison

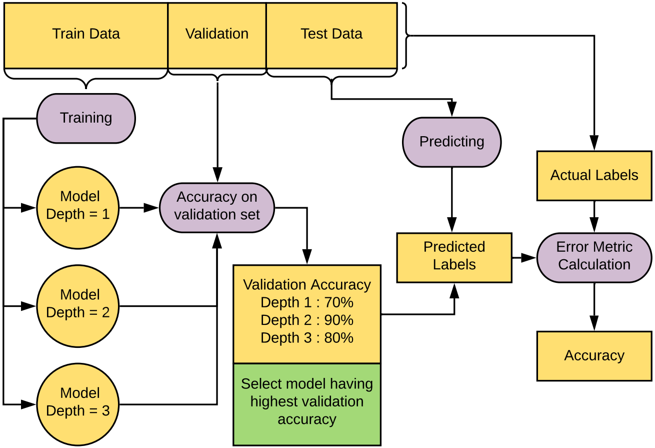
Our General Training Flow



K-fold CV: utilize full dataset



Validation Set Workflow



Nested Cross-Validation (schematic)

Cross-Validation

Nipun Batra and teaching staff

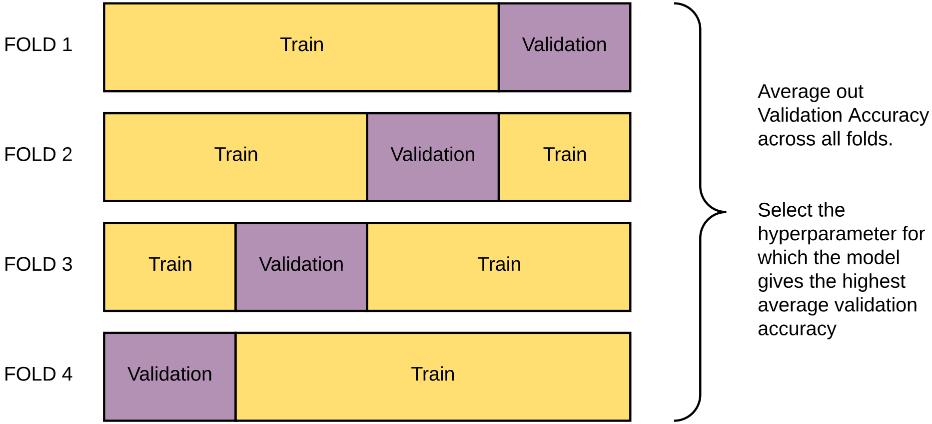
IIT Gandhinagar

August 2, 2025

1 / 22

Nested CV: Model Selection

Average validation performance across folds; pick hyperparameters with best mean (and consider variance).



Roadmap

Pop Quiz

1. What drives high bias? (example)
2. What drives high variance? (example)
3. How does cross-validation help model selection?
4. Why not optimize on test error?

Key Takeaways

- $\text{Error} = \text{Bias}^2 + \text{Variance} + \text{Noise}$

Key Takeaways

- Error $=$ Bias^2 $+$ Variance $+$ Noise
- High Bias $\Rightarrow$ underfit; High Variance $\Rightarrow$ overfit

Key Takeaways

- Error $=$ Bias^2 $+$ Variance $+$ Noise
- High Bias $\Rightarrow$ underfit; High Variance $\Rightarrow$ overfit
- Cross-validation $\Rightarrow$ honest selection

Key Takeaways

- Error $=$ Bias^2 $+$ Variance $+$ Noise
- High Bias $\Rightarrow$ underfit; High Variance $\Rightarrow$ overfit
- Cross-validation $\Rightarrow$ honest selection
- Choose capacity at the *sweet spot*

Next time: Ensembles

- Combining models
- Reducing bias vs reducing variance